View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

JENA ECONOMIC RESEARCH PAPERS

2010 – 082

Should I remember more than you?
- On the best response to factor-based strategies -

by

René Levínský
Abraham Neyman
Miroslav Zelený

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Jena Economic Research Papers 2010 - 082

Should I remember more than you?

– On the best response to factor-based strategies –

René Levínský*
Max Planck Institute of Economics,
Kahlaische Strasse 10, 07745 Jena, Germany,

Abraham Neyman†
Institute of Mathematics and Center for the Study of Rationality,
The Hebrew University of Jerusalem
Giv'at Ram, Jerusalem 91904, Israel,

and Miroslav Zelený‡
Department of Mathematical Analysis,
Faculty of Mathematics and Physics, Charles University,
Sokolovská 83, 186 75 Praha, Czech Republic.

November 29, 2010

JEL classification: C73

Keywords: Bounded rationality, factor-based strategies, bounded recall strategies, finite automata.

*Corresponding author. Email: levinsky@econ.mpg.de

†The second author was supported in part by Israel Science Foundation grant 1123/06.

‡The third author was supported by the research project MSM 0021620839 financed by MSMT.

Abstract

In this paper we offer a new approach to modeling strategies of bounded complexity, the so-called factor-based strategies. In our model, the strategy of a player in the multi-stage game does not directly map the set of histories H to the set of her actions. Instead, the player's perception of H is represented by a factor $\varphi : H \to X$, where X reflects the "cognitive complexity" of the player. Formally, mapping φ sends each history to an element of a factor space X that represents its equivalence class. The play of the player can then be conditioned just on the elements of the set X.

From the perspective of the original multi-stage game we say that a function φ from H to X is a factor of a strategy σ if there exists a function ω from X to the set of actions of the player such that $\sigma = \omega \circ \varphi$. In this case we say that the strategy σ is φ-factor-based. Stationary strategies and strategies played by finite automata and strategies with bounded recall are the most prominent examples of factor-based strategies.

In the discounted infinitely repeated game with perfect monitoring, a best reply to a profile of φ-factor-based strategies need not be a φ-factor-based strategy. However, if the factor φ is recursive, namely its value $\varphi(a_1, \ldots, a_t)$ on a finite string of action profiles $(a_1, \ldots, a_t)$ is a function of $\varphi(a_1, \ldots, a_{t-1})$ and a_t, then for every profile of factor-based strategies there is a best reply that is a pure factor-based strategy.

We also study factor-based strategies in the more general case of stochastic games.

1. Introduction

There are two widely studied approaches to modeling strategies of bounded complexity in (infinitely) repeated games. Aumann (1981), Lehrer (1988), and Aumann and Sorin (1989) consider players with stationary bounded recall strategies (SBR strategies) who have imperfect consciousness of the actual stage of the game, and whose action in the current stage game relies only on the t previous signals they observed and can "remember." Neyman (1985), Rubinstein (1986), Abreu and Rubinstein (1988), and Ben-Porath (1993) deal with (infinitely) repeated games in which players are represented by finite automata (Moore machines). Both models provide a measure of the complexity of the strategy. In the bounded recall approach, the complexity of a strategy is described by the "depth of recall" t, and the complexity of a strategy played by an automaton is measured by the minimal number of states the automaton must have to play the given strategy.

In this paper we pursue the question already raised by Kalai (1990): "What information system (size and structure) should a player maintain when playing a strategic game?" in the context of strategies of bounded complexity. In detail, we study the complexity of the strategy that is the best response to a strategy with a given complexity. Abreu and Rubinstein (1988) show that for every finite automaton A^1 in the discounted repeated game, there exists a finite automaton A^2 such that A^2 maximizes its own payoff in the game against A^1 and the number of states of A^2 is less than or equal to the number of states of A^1. Here, we address this question in the broader context of the newly defined concept of factor-based strategies.

In our bounded rationality approach, the player is not cognitively capable of processing the set of all possible strategies as the set of all possible mappings from the set of all (finite) histories H to the set of actions. Instead, the player can base her actions only on elements x from some abstract set X, where the set X reflects the set of histories H through a mapping $\varphi : H \to X$. Here, φ describes the player's capacity to differentiate between elements of

H. Alternatively, we can understand X as an image (a representation) of H in the players mind, where an element x of X represents the set of histories $H_x = \{h \in H : \varphi(h) = x\}$. Naturally, we are interested in cases where the set X is a proper factor of H.

In defining the factor-based strategies, we were originally motivated by bounded recall strategies. The player is unable to distinguish between two different histories h and h' in the case where the two histories are identical in the last t coordinates. This fact can be easily described by $\varphi(h) = \varphi(h') = x$. Our formal approach can capture much more than SBR strategies. The strategies played by finite automata are also factor-based (with finite range X).

Moreover, we can easily "translate" our model to Aumann (1976); the state space Ω corresponds to the set of histories H, and the partition $\mathcal{P}$ is defined by $\mathcal{P} = \{H_x : x \in X\}$. Here we can easily see that the factor-based strategies can model a player whose cognitive failure is of a different nature than forgetfulness; e.g., a player with infinite recall who is unable to distinguish between some actions of her opponent (i.e., games with imperfect monitoring). Again, the strategies of such a player will be factor-based (possibly with infinite range X).

The concept of an agent with limited ability to distinguish between histories reflects also an older invention: the modal frame $\langle W, R\rangle$ of Kripke (1959). Here, the elements of W represent the "possible worlds" and the binary relation R on W is known as the *accessibility relation*. Identifying W with the set of histories H, and R with an equivalence relation, we match the concept of factor-based strategies with the structure of a modal frame.

With the concept of factor-based strategies in hand, we can come back to the original question "what is the complexity of the strategy that is the best response to a strategy with a given complexity?" In our model this means the following: Consider player 1 endowed with the set of actions A_1 who "lives" in "mental world" $\varphi : H \to X$, and plays some strategy $\omega^1 \circ \varphi$, where $\omega^1 : X \to A_1$. Now consider a (general, unbounded) strategy σ^2 that is the best response strategy of player 2 to $\omega^1 \circ \varphi$ and another strategy $\omega^2 \circ \varphi$ that is the

best response to $\omega^1 \circ \varphi$ from the class of the "bounded" φ-based strategies. Now we ask ourselves under which circumstances does σ^2 fare better than $\omega^2 \circ \varphi$ against $\omega^1 \circ \varphi$. In other words: considering the mental model of my opponent represented by $\varphi : H \to X$, under which conditions on φ is it really profitable for me to be "cleverer" than my opponent (i.e., to play with a general σ^2 that is the mapping from the whole set of histories H), and when is it enough to be just as "clever" as she is (i.e., to play just using some ω^2 that maps only X to the set of my actions).

As the first (negative) result of our paper we show that in the discounted infinitely repeated game with perfect monitoring, a best reply to a profile of φ-factor-based strategies need not be a φ-factor-based strategy. We obtain our main (positive) result for φ that is recursive, i.e., if there exists a function $g : X \times A \to X$ such that $\varphi(a_1, \ldots, a_t) = g(\varphi(a_1, \ldots, a_{t-1}), a_t)$, where A is the set of action profiles in the stage game. Note that in all the examples of factor-based strategies above (finite automata, SBR strategies, imperfect monitoring) the factor φ is recursive. For every recursive factor φ we show that for any profile of factor-based strategies there is a best reply that is a pure factor-based strategy.

As a tool we use the theory of Markov decision processes (MDP), namely theorems on the existence of the best stationary strategy for a given MDP. In fact, once we rephrase our problem of finding the best reply as a question in an MDP our results turn out to be corollaries of the results of Blackwell (1962) and Derman (1965).

This new perspective on Blackwell's optimality also proves (and extends) the previous results of Abreu and Rubinstein (1988). First, the statements are now proven in the same way for behavioral automata and behavioral SBR strategies. Second, Blackwell's theorem gives all statements in a more robust form for patient players, namely for the whole interval of discount factors $\beta \in [\beta_0, 1)$.

All relevant notions will be defined and discussed in the next section. Section 3 introduces the concept of factor-based strategies and presents examples. Section 4 contains the main result and its proof. Section 5 concludes.

2. The Game Models

If X is a finite or countable set (or a measurable space), then $\Delta(X)$ denotes the set of all probabilities on X. Our results apply to a large class of multistage games with perfect monitoring.

2.1. **Supergames.** We start with recalling the model of the two-person supergame with finite action sets. Let $G = \langle A_1, A_2, u_1, u_2\rangle$ be a stage game, where A_i is a nonempty finite set of actions for player i $(i = 1, 2)$ and $u_i : A_1 \times A_2 \to \mathbb{R}$ is the payoff function of player i. The corresponding supergame G^∞ is played as follows. At each period $t \in \mathbb{N} = \{1, 2, 3, \dots\}$ players 1 and 2 make simultaneous and independent moves $a_t^i \in A_i$, $i = 1, 2$.

A play of the supergame is a sequence of action profiles $(a_t)_{t=1}^\infty$ with $a_t = (a_t^1, a_t^2) \in A = A_1 \times A_2$, and a play $(a_t)_{t=1}^\infty$ defines a stream $(u_i(a_t))_{t=1}^\infty$ of payoffs to player i.

A *pure strategy for player i in the supergame G^∞* is a mapping $\sigma : A^{<\mathbb{N}} \to A_i$. The player i following a pure strategy σ plays at the t-th round the action $\sigma(a_1, \dots, a_{t-1})$ where $(a_1, \dots, a_{t-1}) \in A^{t-1}$ is the sequence of actions that have been already played.

A *behavioral strategy for player i in the supergame G^∞* is a mapping $\sigma : A^{<\mathbb{N}} \to \Delta(A_i)$. Player i following a behavioral strategy σ plays at the t-th round an action $a_t^i \in A_i$ with the probability $\sigma(a_1, \dots, a_{t-1})(a_t^i)$ where $(a_1, \dots, a_{t-1}) \in A^{t-1}$ is the sequence of actions that have been already played. Pure strategies can be viewed as a special case of behavioral strategies by identifying A_i with the Dirac measures on A_i. This point of view will be used throughout the paper.

2.2. **Supergames with a time-dependent stage game.** The previous concept can be generalized as follows. Let $\{\langle A_1(t), A_2(t), u_1(t), u_2(t)\rangle\}$ be a sequence of stage games. The corresponding game Γ^∞ is played as follows. At each period $t \in \mathbb{N}$ players 1 and 2 make simultaneous and independent moves $a_t^i \in A_i(t)$, $i = 1, 2$. These plays define a stream $(u_i(t)(a_t))_{t=1}^\infty$ of payoffs to player i. The pure and behavioral strategies of player i in Γ^∞ are defined in a straightforward way.

2.3. **Stochastic games.** A two-person stochastic game with finite action sets is 5-tuple $\Gamma = \langle S, A, u, p, \mu \rangle$ such that

- a *state space* S is a nonempty set,
- $A(z) = A_1(z) \times A_2(z)$ is an *action set*: for every state $z \in S$, $A_i(z)$ is a nonempty finite set of actions for player i $(i = 1, 2)$ at the state z,
- $u = (u_1, u_2)$ is a *payoff function*, where $u_i(z, a)$ is the payoff function of player i, $(z \in S, a \in A(z))$,
- p is a *transition function*: for each state $z \in S$ and each action profile $a \in A(z)$, $p(z, a) \in \Delta(S)$ is a probability distribution of next states; i.e., $p(z, a)(z')$ is the probability of moving to the state z' if the players played a at the state z, and
- $\mu \in \Delta(S)$ is a distribution of the initial state.

A play of the stochastic game Γ^∞ is a sequence of states and actions $(z_1, a_1, \ldots, z_t, a_t, z_{t+1}, a_{t+1}, \ldots)$ with $a_t \in A(z_t)$.

A pure strategy of player i in the stochastic game with perfect monitoring specifies her action $a_t^i \in A_i(z_t)$ as a function of the past state and action profiles $(z_1, a_1, \ldots, a_{t-1}, z_t)$. Similarly, a behavioral strategy of player i is a function of the past state and action profiles $(z_1, a_1, \ldots, a_{t-1}, z_t)$ and specifies the probability that an action $a_t^i \in A_i(z_t)$ is played. A pair of strategies σ^1 and σ^2 of players 1 and 2 defines a probability distribution P_{σ^1,σ^2} on the space of plays of the stochastic game. The expectation w.r.t. this probability distribution is denoted by E_{σ^1,σ^2}. Given a discount factor $0 < \beta < 1$ the (unnormalized) β-discounted payoff to player i is defined by

$$V_\beta^i(\sigma^1, \sigma^2) = E_{\sigma^1,\sigma^2}\left(\sum_{t=1}^{\infty} \beta^{t-1} u_i(z_t, a_t)\right)$$

and the *normalized β-discounted payoff* to player i is defined by

$$v_\beta^i(\sigma^1, \sigma^2) = (1 - \beta) V_\beta^i(\sigma^1, \sigma^2).$$

This normalization ensures that if player i receives a payoff c at each period (i.e., the stream of her payoffs is constant), then $v^i_\beta(\sigma^1, \sigma^2) = c$.

Supergames are a special case of stochastic games with a single state. Similarly, supergames with a time-dependent stage game can be viewed as stochastic games with the state space $\mathbb{N}$ and the deterministic transition $t \mapsto t+1$. Thus, the normalized β-discounted payoff is well defined also for supergames (possibly with a time-dependent stage game) as long as their stage payoffs are either bounded or grow in a subexponential rate in t. Therefore, results on stochastic games will have direct consequences for them.

3. Factor-based strategies

Let H denote the set of all finite histories in a supergame G^∞ (in a stochastic game respectively), i.e., $H = A^{<\mathbb{N}}$ ($H = S \times (A \times S)^{<\mathbb{N}}$ respectively). Let X be a set and φ be a mapping from H to X.

We say that a behavioral strategy σ is a *factor-based strategy with factor* φ (φ-based strategy for short) for player i in the supergame G^∞ if there is a *factor-action function* $\omega : X \to \Delta(A_i)$ such that $\sigma = \omega \circ \varphi$. Factor φ is called *recursive* if there is a function $g : X \times A \to X$ such that $\varphi(a_1, \ldots, a_t) = g(\varphi(a_1, \ldots, a_{t-1}), a_t)$.

The notion of *factor-based strategy* for player i in the supergame Γ^∞ with a time-dependent stage game is defined analogously. The resulting probability of a^i_t depends on $\varphi(a_1, \ldots, a_{t-1})$ and on the actual period t. Thus the φ-based strategy σ satisfies

$$\sigma(a_1, a_2, \ldots, a_{t-1}) = \omega(t, \varphi(a_1, a_2, \ldots, a_{t-1})).$$

for some $\omega : \mathbb{N} \times X \to \Delta(A_i)$.

Further, we define *φ-based strategy* for player i in the stochastic game. The choice of distribution of action a^i_t depends on $\varphi(z_1, a_1, \ldots, z_{t-1}, a_{t-1})$ and on the actual state z_t. This means that $\omega : S \times X \to \Delta(A_i)$ and

$$\sigma(z_1, a_1, \ldots, z_t) = \omega(z_t, \varphi(z_1, a_1, \ldots, z_{t-1}, a_{t-1})).$$

Factor φ in the case of a stochastic game is called *recursive* if there is a function $g : X \times S \times A \to X$ such that $\varphi(z_1, a_1, \ldots, z_t, a_t) = g(\varphi(z_1, a_1, \ldots, z_{t-1}, a_{t-1}), z_t, a_t)$. A few classes of recursive φ-based strategies follow.

3.1. **SBR strategies.** Let $k \in \mathbb{N}$. By a *behavioral k-SBR strategy for player i* in the supergame G^∞ we mean a pair (e, ω), where $e = (e_1, e_2, \ldots, e_k) \in A^k$ and $\omega : A^k \to \Delta(A_i)$ is a mapping. Player i following the strategy (e, ω) plays as follows. If moves $a_1, \ldots, a_l \in A$ have been played, then player i takes the sequence s, which is formed by the last k elements of the sequence $(e_1, \ldots, e_k, a_1, \ldots, a_l)$, and his $(l+1)$-th move is $a \in A_i$ with the (conditional) probability $\omega(s)(a)$. A *pure k-SBR strategy for player i* in the supergame G^∞ is defined in a straightforward way.

Defining $\varphi(a_1, \ldots, a_l)$ to be the last k elements of the sequence $(e_1, \ldots, e_k, a_1, \ldots, a_l)$, the k-SBR strategy σ defined above obeys $\sigma = \omega \circ \varphi$, and φ is recursive; thus σ is a recursive φ-based strategy with finite range.

We say that a behavioral (pure) strategy σ is a *behavioral (pure) SBR strategy* if σ is a behavioral (pure) k-SBR strategy for some $k \in \mathbb{N}$.

3.2. **Strategies with time-dependent recall.** (See, e.g., Neyman and Okada, 2009.) Let $k : \mathbb{N} \to \mathbb{N}$ be a function with $k(t) < t$ for every $t \in \mathbb{N}$. *Behavioral* (respectively, *pure*) *$k(t)$-BR strategy* is defined analogously to the above case but the action at stage t depends on t and the last $k(t)$ stage-actions. Let σ be such a strategy. Setting $\varphi(a_1, \ldots, a_t) = (t, (a_{t-k(t)}, \ldots, a_{t-1}))$ we easily see that σ is φ-based. Moreover, φ is recursive provided $k(t+1) \leq k(t) + 1$ for every $t \in \mathbb{N}$.

3.3. **Automata and behavioral automata.** A *behavioral automaton* (for player 1 in the supergame G^∞) is a quadruple $\langle M, m^*, \alpha, \tau \rangle$, where M is a nonempty set (the state space), $m^* \in M$ is the initial state, $\alpha : M \to \Delta(A_1)$ is a probabilistic action function, and $\tau : M \times A \to M$ is a transition function. A *k-state behavioral automaton* is a behavioral automaton where the set M has k elements. A behavioral automaton $\langle M, m^*, \alpha, \tau \rangle$ defines a behavioral

strategy σ^1 (for player 1) inductively: $m_1 = m^*$, $\sigma^1(\emptyset) = \alpha(m_1)$, $\sigma^1(a_1, \ldots, a_{t-1}) = \alpha(m_t)$, where $m_t = \tau(m_{t-1}, a_{t-1})$.

A behavioral automaton $\langle M, m^*, \alpha, \tau \rangle$ defines a recursive φ-based strategy where $X = M$, $\varphi(\emptyset) = m^*$, $\varphi(a_1, \ldots, a_t) = \tau(\varphi(a_1, \ldots, a_{t-1}), a_t)$, and $\omega = \alpha$.

A *k-state (deterministic) automaton* is defined by the replacement of $\Delta(A_1)$ with A_1.

3.4. **Time-dependent automata.** (See, e.g., Neyman, 1997.) A time-dependent action automaton is defined by replacing the action function α by a sequence of action functions α_t, $t \geq 1$, where α_t defines the action at stage t. Similarly, a time-dependent transition automaton is obtained by replacing the (stationary) transition function τ with a sequence of time-dependent transitions τ_t, $t \geq 1$, where τ_t defines the transition at stage t. Finally, a time-dependent (action and transition) automaton in the supergame G^∞ is a quadruple $\langle M, m^*, (\alpha_t)_{t=1}^\infty, (\tau_t)_{t=1}^\infty \rangle$, where M is a nonempty set (the state space), $m^* \in M$ is the initial state, $\alpha_t : M \to \Delta(A_1)$ is a probabilistic action function, and $\tau_t : M \times A \to M$ is a (deterministic) transition function. It defines a behavioral strategy σ^1 (for player 1) inductively: $m_1 = m^*$, $\sigma^1(\emptyset) = \alpha_1(m_1)$, $\sigma^1(a_1, \ldots, a_{t-1}) = \alpha_t(m_t)$, where $m_t = \tau_t(m_{t-1}, a_{t-1})$. Note that a time-dependent automaton $\langle M, m^*, (\alpha_t)_{t=1}^\infty, (\tau_t)_{t=1}^\infty \rangle$ defines the same strategy as the automaton $\langle M \times \mathbb{N}, m^{**}, \alpha, \tau \rangle$ with $m^{**} = (m^*, 1)$, $\alpha(m, t) = \alpha_t(m)$ and $\tau((m, t), a) = (\tau_t(m, a), t + 1)$. Therefore, the corresponding strategy is a recursive φ-based strategy, where $\varphi : A^{<\mathbb{N}} \to M \times \mathbb{N}$ is given by $\varphi(a_1, \ldots, a_t) = \tau(\varphi(a_1, \ldots, a_{t-1}), a_t)$ and $\omega = \alpha$.

3.5. **A counterexample.** Our objective is to study for what factors φ of the strategy σ^1 player 2 has a φ-based best reply. First, we demonstrate that in the discounted two-person repeated game (with finitely many stage actions) there need not be such a strategy.

Let G be the stage game with stage-action sets $A_1 = A_2 = \{1, 2\}$, and the payoff function to player 2 is given by $u_2(1, 1) = u_2(1, 2) = 0$, $u_2(2, 1) = u_2(2, 2) = 1$. Define the factor

$\varphi : H \to X$ by $X = \{B, C\}$ and

$$\varphi(h) = \begin{cases} B & \text{if } \big(h = (a_1) \text{ and } a_1^2 = 1\big) \text{ or } \big(h = (a_1, a_2, a_3) \text{ and } a_3^2 = 2\big), \\ C & \text{otherwise.} \end{cases}$$

Consider a φ-based strategy σ^1 defined via $\omega^1 : X \to A_1$, where $\omega^1(B) = 2$ and $\omega^1(C) = 1$. Let us demonstrate that any φ-based strategy σ^2 of player 2 cannot be a best reply to the strategy σ^1. First, the nonzero payoffs to player 2 are possible only in stages 2 and 4. Suppose σ^2 is φ-based with $\sigma^2 = \omega^2 \circ \varphi$. Set $0 \le \omega^2(C)(1) = x \le 1$. Then $V_\beta^2(\sigma^1, \sigma^2) = \beta x + \beta^3(1 - x)$. But the strategy $\tilde{\sigma}^2$, where player 2 plays 1 in the first period and 2 in the third, yields $V_\beta^2(\sigma^1, \tilde{\sigma}^2) = \beta + \beta^3 > \beta x + \beta^3(1 - x)$, whenever $x \in [0, 1], \beta \in (0, 1)$.

4. Main results

The main result follows.

Theorem 4.1. *Let $\Gamma = \langle S, A, u, p, \mu \rangle$ be a two-person stochastic game with countably many states, finitely many actions at each state, and a bounded payoff function u_2. Let σ^1 be a φ-based behavioral strategy of player 1 in Γ^∞. If φ is recursive, then the following hold.*

(i) *For every $\beta \in (0, 1)$ there exists a φ-based pure strategy σ^2 such that for every behavioral strategy ρ of player 2 in Γ^∞ we have $v_\beta^2(\sigma^1, \sigma^2) \ge v_\beta^2(\sigma^1, \rho)$.*

(ii) *If S and the range of φ are, in addition, finite, then there is a φ-based pure strategy σ^2 and a discount factor $\beta_0 \in (0, 1)$ such that*

- *for every behavioral strategy ρ (of player 2 in Γ^∞) and every $\beta \in [\beta_0, 1)$, we have $v_\beta^2(\sigma^1, \sigma^2) \ge v_\beta^2(\sigma^1, \rho)$;*
- *for every behavioral strategy ρ, we have*

$$E_{\sigma^1, \sigma^2}\left(\liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} u_2(z_t, a_t)\right) \ge E_{\sigma^1, \rho}\left(\limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} u_2(z_t, a_t)\right);$$

– *for every* $\varepsilon > 0$, *there exists* $N \in \mathbb{N}$ *such that, for every behavioral strategy* ρ *and every* $n \geq N$, *we have*

$$E_{\sigma^1,\sigma^2}\left(\frac{1}{n}\sum_{t=1}^{n} u_2(z_t, a_t)\right) \geq E_{\sigma^1,\rho}\left(\frac{1}{n}\sum_{t=1}^{n} u_2(z_t, a_t)\right) - \varepsilon.$$

Remark 4.2. (i) Let G^∞ be a supergame (supergame with time-dependent stage game respectively). Since such a supergame belongs also to the class of stochastic supergames, Theorem 4.1 gives the following consequences in the β-discounted game G^∞, $\beta \in (0, 1)$.

a) For every behavioral k-SBR strategy σ^1, there is a pure k-SBR strategy σ^2 that is a best reply of player 2.

b) For every behavioral (time-dependent recall) $k(t)$-SBR strategy σ^1 with $k(t+1) \leq k(t) + 1$, there is a pure $k(t)$-SBR strategy σ^2 that is a best reply.

c) For every strategy σ^1 that is defined by a k-state behavioral automaton, there is a best reply σ^2 of player 2 defined by a (deterministic) k-state automaton.

d) For every strategy σ^1 that is defined by a k-state time-dependent automaton, there is a best reply σ^2 defined by a (deterministic) k-state time-dependent automaton.

(ii) The extension of the models from two-person games to multi-person games is straightforward and our results on the best reply for two-person games extends to n-person games ($n > 2$) since players $1, 2, \ldots, n-1$ can be considered as one player playing actions from the space $A_1 \times \cdots \times A_{n-1}$.

The main result is a simple corollary of results on Markov decision processes. By a *Markov decision process* (*MDP*, for short) we mean a one-person (called the decision maker) stochastic game. We recall the definition of the MDP in the following notation: by r we denote the single-stage payoff function to the decision maker and by $v_\beta(\sigma)$ the normalized β-discounted payoff to the decision maker when the strategy σ is played.

More precisely, by MDP we mean a 5-tuple $\langle M, B, r, p, \nu \rangle$ such that

- M is a nonempty countable set (set of states),
- $B(z), z \in M$ is a nonempty finite set (set of actions at the state z),
- $r(z, a)$ is a real number for every $z \in M$ and $a \in B(z)$ (reward function),
- $p(z, a)$ is a probability on M for every $z \in M$ and $a \in B(z)$,
- ν is an initial probability on M.

One can interpret this structure as follows. The set $B(z)$ is the set of feasible actions that can be played at state $z \in M$ by the decision maker. The sequence $(z_1, a_1, z_2, a_2, \dots)$ of states and actions of the process is realized as follows. The initial z_1 is chosen with the probability $\nu(z_1)$. If the sequence $(z_1, a_1, z_2, a_2, \dots, z_t)$ has been constructed, then the decision maker plays an action $a_t \in B(z_t)$ and receives a payoff $r(z_t, a_t)$. The (conditional) probability of the next state $z_{t+1} \in M$ of the process (given $z_1, \dots, z_t, a_t$) is given by the probability distribution $p(z_t, a_t)$.

A *strategy* for an MDP is a function σ that assigns to every finite sequence of states and actions $s = (z_1, a_1, z_2, a_2, \dots, z_t)$ a probability $\sigma(s)$ on $B(z_t)$. If $\sigma(s)$ is always a Dirac measure, then σ is *pure*. By a *stationary strategy* for an MDP we mean a strategy depending only on the last state.

A strategy σ of the decision maker defines a probability distribution P_σ on the space of plays of the MDP. The expectation w.r.t. this probability distribution is denoted by E_σ. Given a discount factor $0 < \beta < 1$, the normalized β-discounted payoff to the decision maker is defined by

$$v_\beta(\sigma) = (1 - \beta) \cdot E_\sigma \left(\sum_{t=1}^{\infty} \beta^{t-1} r(z_t, a_t) \right).$$

The key tools in our paper are results of Blackwell and Derman. Parts (ii) and (iii) of the Theorem 4.4 follow implicitly from part (i) and the proof in Mertens and Neyman (1981) that shows that the stationary strategy σ that obeys (i) is ε-optimal for every $\varepsilon > 0$; for an explicit statement see Neyman (2003).

Theorem 4.3 (Derman, 1965)**.** *Let $\langle M, B, r, p, \nu\rangle$ be an MDP with countably many states and finitely many actions in each state, and with bounded reward function. Then for each $\beta \in (0,1)$ there is a stationary pure strategy σ such that, for every strategy ρ, we have $v_\beta(\sigma) \geq v_\beta(\rho)$.*

Theorem 4.4 (Blackwell, 1962)**.** *Let $\langle M, B, r, p, \nu\rangle$ be an MDP with finitely many states and actions. Then there is a stationary pure strategy σ and a discount factor $\beta_0 \in (0,1)$ such that*

(i) *for every strategy ρ and for every $\beta \in [\beta_0, 1)$, we have $v_\beta(\sigma) \geq v_\beta(\rho)$;*

(ii) *for every strategy ρ, we have*

$$E_\sigma\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r(z_t, a_t)\right) \geq E_\rho\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r(z_t, a_t)\right);$$

(iii) *for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that, for every strategy ρ and every $n \geq N$, we have*

$$E_\sigma\left(\frac{1}{n}\sum_{t=1}^{n} r(z_t, a_t)\right) \geq E_\rho\left(\frac{1}{n}\sum_{t=1}^{n} r(z_t, a_t)\right) - \varepsilon.$$

Proof of Theorem 4.1. Let σ^1 be a φ-based strategy for player 1 in the stochastic game Γ and assume that φ is recursive. Thus, there exist functions $\omega : S \times X \to \Delta(A_1)$ and $g : X \times S \times A \to X$ such that

$$\sigma^1(z_1, a_1, \ldots, z_t) = \omega(z_t, \varphi(z_1, a_1, \ldots, z_{t-1}, a_{t-1})),$$
$$\varphi(z_1, a_1, \ldots, z_t, a_t) = g(\varphi(z_1, a_1, \ldots, z_{t-1}, a_{t-1}), z_t, a_t).$$

We define an MDP $\mathfrak{M} = \langle M, B, r, q, \nu \rangle$ as follows.

$$
\begin{aligned}
M &= S \times X, \\
B(z,x) &= A_2(z), \qquad (z,x) \in M, \\
r(z,x,a^2) &= \sum_{a^1 \in A_1(z)} u_2(z,(a^1,a^2)) \cdot \omega(z,x)(a^1), \qquad (z,x) \in M,\ a^2 \in A_2(z), \\
q(z,x,a^2)(z',x') &= \sum_{\substack{a^1 \in A_1(z), \\ g(x,z,(a^1,a^2))=x'}} p(z,(a^1,a^2))(z') \cdot \omega(z,x)(a^1), \qquad (z',x') \in M,\ a^2 \in A_2(z), \\
\nu(z,x) &= \begin{cases} \mu(z), & x = \varphi(\emptyset); \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

A play of $\mathfrak{M}$ is of the form

$$(z_1, x_1, a_1^2, z_2, x_2, a_2^2, \dots, z_t, x_t, a_t^2, \dots).$$

If ρ is a strategy for player 2 in Γ^∞, then the probability measure $P_{\sigma^1,\rho}$ captures probability distribution of possible plays $(z_1, a_1, z_2, \dots)$ of Γ^∞, where players 1 and 2 follow the strategies σ^1 and ρ, respectively. Thus, $P_{\sigma^1,\rho}(z_1, a_1, \dots, z_t)$ is the probability that a play starts with the sequence $(z_1, a_1, \dots, z_t)$. Similarly, if ζ is a strategy of the decision maker in $\mathfrak{M}$, then the probability that a play starts with $(z_1, x_1, a_1^2, \dots, z_t, x_t)$ is denoted by $P_\zeta(z_1, x_1, a_1^2, \dots, z_t, x_t)$.

Let ψ be a mapping assigning to each sequence $(z_1, a_1, \dots, z_t)$ the corresponding sequence $(z_1, x_1, a_1^2, \dots, z_t, x_t)$, where $x_1 = \varphi(\emptyset)$, $x_j = g(x_{j-1}, z_{j-1}, a_{j-1})$, $j = 2, \dots, t$.

Let ρ be a strategy of player 2 in Γ^∞. Then we define the corresponding strategy $\tilde{\rho}$ in $\mathfrak{M}$ by

$$\tilde{\rho}(\tilde{s})(a^2) = \sum_{s, \psi(s)=\tilde{s}} \rho(s)(a^2) \cdot P_{\sigma^1,\rho}(s|\tilde{s}),$$

where $s = (z_1, a_1, \dots, z_t)$, $\tilde{s} = (z_1, x_1, a_1^2, \dots, z_t, x_t)$, and $P_{\sigma^1,\rho}(s|\tilde{s})$ denotes the conditional probability of s given $\tilde{s}$. The symbol $P_{\sigma^1,\rho}(\tilde{s})$ denotes the probability that the play starts

with a sequence s satisfying $\psi(s) = \tilde{s}$, that is,

$$P_{\sigma^1,\rho}(\tilde{s}) = \sum_{s,\psi(s)=\tilde{s}} P_{\sigma^1,\rho}(s).$$

Claim 4.5.

(i) *For every fixed* $\tilde{s} = (z_1, x_1, a_1^2, \ldots, z_t, x_t)$ *we have* $P_{\sigma^1,\rho}(\tilde{s}) = P_{\tilde{\rho}}(\tilde{s})$.

(ii) *Let* $\beta \in (0,1)$, $t \in \mathbb{N}$, *and* ρ *be a strategy in* Γ^∞ *for player 2. Then* $E_{\tilde{\rho}}(r(z_t, x_t, a_t^2)) = E_{\sigma^1,\rho}(u_2(z_t, a_t))$.

Proof of Claim. (i) We will proceed by induction on the length of $\tilde{s}$. Suppose that $\tilde{s} = (z_1, x_1)$. If $x_1 = \varphi(\emptyset)$, then we clearly have

$$\sum_{s,\psi(s)=\tilde{s}} P_{\sigma^1,\rho}(s) = P_{\sigma^1,\rho}(z_1) = \mu(z_1) = P_{\tilde{\rho}}(z_1, x_1).$$

If $x_1 \neq \varphi(\emptyset)$, then the equality clearly holds. Now assume that the desired equality holds for every $\tilde{w} = (z_1, x_1, a_1^2, \ldots, z_t, x_t)$. Fix such a $\tilde{w}$ and consider $\tilde{s}$ of the form $\tilde{s} = (z_1, x_1, a_1^2, \ldots, z_t, x_t, a_t^2, z_{t+1}, x_{t+1})$. We have

$$\begin{aligned}
\sum_{s,\psi(s)=\tilde{s}} P_{\sigma^1,\rho}(s) &= \sum_{\substack{w \\ \psi(w)=\tilde{w}}} \sum_{\substack{a^1 \in A_1(z_t) \\ g(x_t,z_t,(a^1,a_t^2))=x_{t+1}}} p(z_t,(a^1,a_t^2))(z_{t+1}) \cdot \omega(z_t,x_t)(a^1) \cdot P_{\sigma^1,\rho}(w) \\
&= \sum_{\substack{w \\ \psi(w)=\tilde{w}}} q(z_t, x_t, a_t^2)(z_{t+1}, x_{t+1}) \cdot P_{\sigma^1,\rho}(w) \\
&= q(z_t, x_t, a_t^2)(z_{t+1}, x_{t+1}) \cdot P_{\tilde{\rho}}(\tilde{w}) \qquad \text{(by induction hypothesis)} \\
&= P_{\tilde{\rho}}(\tilde{s}).
\end{aligned}$$

(ii) Let us compute

$$\begin{aligned}
E_{\sigma^1,\rho}(u_2(z_t, a_t)) &= \int u_2(z_t, a_t)\, dP_{\sigma^1,\rho} \\
&= \sum_{s=(z_1,a_1,\ldots,z_t)} \sum_{a=(a^1,a^2)\in A(z_t)} u_2(z_t, a) \cdot \omega(z_t, x_t)(a^1) \cdot \rho(s)(a^2) \cdot P_{\sigma^1,\rho}(s).
\end{aligned}$$

Using the definition of r we get

$$\begin{aligned}
E_{\sigma^1,\rho}(u_2(z_t,a_t)) &= \sum_{s}\sum_{a^2\in A_2(z_t)} r(z_t,x_t,a^2)\cdot\rho(s)(a^2)\cdot P_{\sigma^1,\rho}(s)\\
&= \sum_{\tilde{s}}\sum_{s,\psi(s)=\tilde{s}}\sum_{a^2\in A_2(z_t)} r(z_t,x_t,a^2)\cdot\rho(s)(a^2)\cdot P_{\sigma^1,\rho}(s)\\
&= \sum_{\tilde{s}}\sum_{a^2\in A_2(z_t)} r(z_t,x_t,a^2)\cdot\left(\sum_{s,\psi(s)=\tilde{s}}\rho(s)(a^2)\cdot P_{\sigma^1,\rho}(s|\tilde{s})\right)\cdot P_{\sigma^1,\rho}(\tilde{s})\\
&= \sum_{\tilde{s}}\sum_{a^2\in A_2(z_t)} r(z_t,x_t,a^2)\cdot\tilde{\rho}(\tilde{s})(a^2)\cdot P_{\sigma^1,\rho}(\tilde{s}).
\end{aligned}$$

Using part (i) of Claim 4.5 we conclude

$$\begin{aligned}
E_{\sigma^1,\rho}(u_2(z_t,a_t)) &= \sum_{\tilde{s}}\sum_{a^2\in A_2(z_t)} r(z_t,x_t,a^2)\cdot\tilde{\rho}(\tilde{s})(a^2)\cdot P_{\tilde{\rho}}(\tilde{s})\\
&= E_{\tilde{\rho}}(r(z_t,x_t,a_t^2)).
\end{aligned}$$

□

Fix $\beta\in(0,1)$. According to Theorem 4.3 there exists a pure stationary strategy τ for the decision maker in $\mathfrak{M}$. Such a strategy defines a φ-based pure strategy σ^2 of player 2 in Γ^∞ as follows:

$$\sigma^2(z_1,a_1,\ldots,z_t)=\tau(z_t,\varphi(z_1,a_1,\ldots,z_{t-1},a_{t-1})).$$

Now assume that we have a strategy ρ of player 2 in Γ^∞. According to Claim 4.5(ii), we have $v^2_\beta(\sigma^1,\rho)=v_\beta(\tilde{\rho})\le v_\beta(\tau)=v^2_\beta(\sigma^1,\sigma^2)$. Thus we get assertion (i). Assertion (ii) follows from Theorem 4.4 and Claim 4.5(ii). □

5. Concluding remarks

5.1. **Compact action spaces.** A natural extension of our model is to consider players with compact action sets A_i. In this extension, there arises a new problem not found in games with finite action profiles, namely the existence of a best reply to a given strategy σ. Consider, for example, the following two-player supergame, where the sets of actions of

each player is the interval $[0,1]$ and the stage-payoff of player 2 is (at any time) given by $u_2(a^1,a^2)=a^1+a^2$. Now, suppose that player 1 plays the 1-SBR strategy given by

$$\sigma^1(a_1,\dots,a_{t-1})=\begin{cases}1, & \text{if } a^2_{t-1}<1 \text{ and } t>1,\\ 0, & \text{otherwise,}\end{cases} \qquad e_1=(0,0).$$

This strategy is recursively factor-based. Indeed, we set $X=\{B,C\}$ and $\varphi(a_1,\dots,a_{t-1})=B$ if $a^2_{t-1}<1$ and $t>1$, $\varphi(a_1,\dots,a_{t-1})=C$ otherwise; $\omega(B)=1$ and $\omega(C)=0$. Then we have $\sigma^1=\omega\circ\varphi$. However, in the β-discounted game there is no φ-based best reply, and any φ-based reply is dominated by (another) φ-based reply.

Of course, there does not exist any general best reply to σ^1. The difficulty stems from the fact that factor φ is not continuous. However, using, e.g., Maitra (1968) one can generalize our results of part (i) of Theorem 4.1.

5.2. **Public vs. private strategies.** Another interpretation of the φ-based strategies is related to the imperfect monitoring literature. Setting X as the set of all possible histories of public signals, we can identify the φ-based strategies with so-called *public strategies* (see, e.g., Radner, Myerson, and Maskin, 1986). In contrast, *a private strategy* (see, e.g., Kandori and Obara, 2006) is a strategy where the current action depends on the history of public signals (i.e., on elements of X), and, in addition, on private signals (e.g., past private actions). Our question at the outset of this paper can then be reformulated as "Considering my opponent is limited to public strategies only, under which conditions can I exploit my (additional) private signal?"; in other words, "Can private strategies fare better than the public strategies against public strategies?" The answer is that one does not profit from the additional private signal since the factor φ is in this situation obviously recursive.

References

Abreu, D., and A. Rubinstein (1988): "The structure of Nash equilibrium in repeated games with finite automata," *Econometrica*, 56(6), 1259–1281.

Aumann, R. J. (1976): "Agreeing to disagree," *Annals of Statistics*, 4, 1236–1239.

——— (1981): "Survey of repeated games," in *Essays in Game Theory and Mathematical Economics in Honour of Oskar Morgenstern*, pp. 11–42. Wissenschaftsverlag, Bibliographisches Institut, Mannheim.

Aumann, R. J., and S. Sorin (1989): "Cooperation and bounded recall," *Games and Economic Behavior*, 1(1), 5–39.

Ben-Porath, E. (1993): "Repeated games with finite automata," *Journal of Economic Theory*, 59, 17–32.

Blackwell, D. (1962): "Discrete dynamic programing," *Annals of Mathematical Statistics*, 33, 719–726.

Derman, C. (1965): "Markovian sequential control processes: Denumerable state spaces," *Journal of Mathematical Analysis and Applications*, 10, 295–302.

Kalai, E. (1990): "Bounded rationality and strategic complexity in repeated games," in *Game Theory and Applications*, ed. by T. Ichiishi, A. Neyman, and Y. Tauman, pp. 131–157. Academic Press, San Diego.

Kandori, M., and I. Obara (2006): "Efficiency in repeated games revisited: The role of private strategies," *Econometrica*, 74(2), 499–519.

Kripke, S. A. (1959): "A completeness theorem in modal logic," *The Journal of Symbolic Logic*, 4, 1–14.

Lehrer, E. (1988): "Repeated games with stationary bounded recall strategies," *Journal of Economic Theory*, 46(1), 130–144.

Maitra, A. (1968): "Discounted dynamic programming on compact metric spaces," *Sankhyā: The Indian Journal of Statistics, Series A*, 30(2), 211–216.

Mertens, J.-F., and A. Neyman (1981): "Stochastic games," *International Journal of Game Theory*, 10, 53–66.

Neyman, A. (1985): "Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma," *Economics Letters*, 19, 227–229.

——— (1997): "Cooperation, repetition, and automata," in *Cooperation: Game Theoretic Approaches*, ed. by S. Hart, and A. Mas-Colell, vol. 155 of *NATO ASI Series F*, pp. 233–255. Springer-Verlag, New York.

——— (2003): "From Markov chains to stochastic games," in *Stochastic Games and Applications*, ed. by A. Neyman, and S. Sorin, vol. 570 of *NATO Science Series C*, pp. 9–25. Kluwer Academic Publishers, Dordrecht.

Neyman, A., and D. Okada (2009): "Growth of strategy sets, entropy, and nonstationary bounded recall," *Games and Economic Behavior*, 66(1), 404–425.

Radner, R., R. Myerson, and E. Maskin (1986): "An example of a repeated partnership game with discounting and with uniformly inefficient equilibria," *Review of Economic Studies*, 53(172), 59–69.

Rubinstein, A. (1986): "Finite automata play the repeated prisoner's dilemma," *Journal of Economic Theory*, 39(1), 83–96.